PROJECT, DESIGN AND MANAGEMENT (PDM)
https://www.mlsjournals.com/Project-Design-Management
ISSN: 2683-1597

(2025) Project, Design and Management 9(1), 71-89. doi.org/10.35992/pdm.v9i1.4073.

CLIENT SEGMENTATION AND DETECTION OF UNUSUAL OPERATIONS CLASSIFIED INTO RISK DEGREES FOR THE PREVENTION OF MONEY LAUNDERING WITH DATA FROM A FINANCIAL INSTITUTION IN MEXICO BY 2023

SEGMENTACIÓN DE CLIENTES Y DETECCIÓN DE OPERACIONES INUSUALES CLASIFICADOS EN GRADOS DE RIESGO PARA LA PREVENCIÓN DE LAVADO DE DINERO CON DATOS DE UNA INSTITUCIÓN FINANCIERA EN MÉXICO A 2023

José Antonio Ortiz Richards[1]
Banco Nacional de Comercio Exterior, Mexico
(jortizrichards@gmail.com) (https://orcid.org/0000-0001-9795-8001)

Manuscript information:

Recibido/Received: 19/11/2024
Revisado/Reviewed: 03/12/2024
Aceptado/Accepted: 10/05/2025

Keywords:
data mining, money laundering prevention, unusual transactions, segmentation methods.

ABSTRACT

The research develops and validates a quantitative method using data from a financial institution in Mexico, with the objective of obtaining a better understanding of the clients, detecting the possible misuse of the institution in crimes related to the integration and dispersion of illicit financial resources, taking the international recommendations established by the Financial Action Task Force (FATF) and with the provisions in Mexico. Data mining techniques are used, as well as instruments to collect, analyze and use information on the inherent and transactional characteristics of customers. A descriptive statistical analysis is presented and, to achieve adequate segmentation, classification methods based on mobile centers and Ward's hierarchical classification are combined, along with factorial methods, which allows for the identification of changes in behavioral patterns of the variables and the analysis of possible unusual operations, explaining the degree of risk associated with each segment. The results offer a classification of medium and high risk, contrasting with the scoring model currently used, which classifies customers as low risk. In addition, this approach facilitates the suspicion of unusual operations by reducing the number of false alerts. One of the contributions this research offers is the incorporation of qualitative variables for segmentation adapted to the context of Mexico, considering best practices in Colombia and the FATF.

Palabras clave:
minería de datos, prevención de lavado de dinero, operaciones

RESUMEN

La investigación desarrolla y valida un método cuantitativo utilizando datos de una institución financiera en México, con el objetivo de obtener un mejor conocimiento de los clientes, detectar el posible uso indebido de la institución en delitos relacionados con la integración y dispersión de recursos financieros ilícitos

[1] Corresponding author.

inusuales, métodos de segmentación.	atendiendo las recomendaciones internacionales establecidas por el Grupo de Acción Financiera Internacional (GAFI) y con las disposiciones en México. Se emplean técnicas de minería de datos, así como instrumentos para recolectar, analizar y utilizar la información sobre las características inherentes y transaccionales de los clientes. Se presenta un análisis estadístico descriptivo y, para lograr una segmentación adecuada, se combinan métodos de clasificación basados en centros móviles y la clasificación jerárquica de Ward, junto con métodos factoriales, lo que permite identificar cambios en patrones de comportamiento de las variables y analizar posibles operaciones inusuales, explicando el grado de riesgo asociado a cada segmento. Los resultados ofrecen una clasificación de riesgo medio y alto, contrastando con el modelo de puntaje actualmente utilizado, que clasifica a los clientes como de riesgo bajo. Además, este enfoque facilita la sospecha de operaciones inusuales reduciendo el número de alertas falsas. Uno de los aportes que ofrece esta investigación es incorporar variables cualitativas para la segmentación adaptada al contexto de México considerando mejores prácticas en Colombia y de GAFI.

Introduction

One of the international priorities is the fight to prevent money laundering and the financing of terrorism because they represent a high risk to peace, security, stability and the economic development of countries. The effect of not being properly managed, on the one hand, is an increase in the number of victims due to crimes committed by individuals or organizations and, on the other hand, according to BASEL (2022) investors and financial institutions may decide to withdraw or not to start a business in a country that is assessed as high risk for money laundering to avoid exposure to non-compliance, operational and reputational risks.

There are several international organizations to combat it, one of which is the FATF, which periodically reviews money laundering techniques, strengthens its standards and monitors countries to ensure that they fully and effectively implement its 40 recommendations. FATF data is a primary source of information on a country's commitment. Mexico partially complies with some recommendations, in accordance with the evaluation process carried out between 2017 and 2018 (GAFILAT,2023).

Another organization is BASEL, which, among other actions, annually measures "the risk of money laundering and terrorist financing (ML/FT) in jurisdictions around the world"(BASEL, 2023, p. 5). Risk is defined as vulnerability to ML/FT and its counteracting capabilities. The index considers 18 indicators, which differ in focus and scope and are classified into five domains. The score is based on data from publicly available sources, such as the FATF, World Bank, World Economic Forum and Transparency International. In 2022 and 2023, Mexico is assessed at medium risk.

The Mexican Government, derived from the international commitments adopted as a member of the FATF, issues for financial institutions, among various laws and regulations, the general provisions referred to in Article 115 of the Law of Credit Institutions to prevent and detect these acts, which establishes that the "Entities, based on the measurement of the Risks they carry out, must classify their Clients in different Risk Grades that allow them to differentiate them significantly" (SHCP, 2022, p. 67).

On the other hand, the National Banking and Securities Commission (CNBV) in its supervisory work identifies several recurring findings, among which the following stand out: "The model of classifying customers by degree of risk does not consider all criteria, factors and inherent or transactional characteristics of customers" (2020, p. 9).

The financial institution develops and uses a point rating model for compliance purposes. Since its application, it presents results with a concentration of low-risk clients, which have heterogeneous characteristics. The limitations of this model are considered to be:

- It is static, so it does not adjust to changes in customer behavior.
- Generates a high number of false positives.

It is highlighted that ML/FT benefits from the use of artificial intelligence as it significantly improves the processing of large volumes of data in real time, facilitating the accuracy of the results to identify atypical operations (FATF, 2021, p. 35; Martinez et. al., 2022, p. 78).

FATF (2021, p. 38) mentions that ML/FT prevention is favored by segmentation, since by forming homogeneous groups of individuals it is possible to find association between them by knowing their usual behavior and from this identify unusual operations, compared to their segment that may be linked to suspicious behavior of crimes associated with ML/FT.

Therefore, the general therefore, the general objective of this article is to use a segmentation method that allows classifying and detecting the unusualness of customer operations considering data from 2022 and 2023 in order to have more elements to know it, classify it in a risk grade and if necessary suspect ML.

The research questions posed are: How could the financial institution in Mexico segment customers to achieve homogeneous groups within and heterogeneous among them; does customer segmentation allow explaining the degree of risk that their inherent and transactional characteristics expose the financial institution in Mexico to be used for money laundering; and could the financial institution in Mexico detect and suspect unusual operations of its customers that represent a risk of money laundering with the support of segmentation?

To do so, we initially collected and analyzed several articles, such as the one published by (Jovel, 2020) who segmented the members of a savings and loan cooperative in Colombia using quantitative variables. It relies on R software, Cross Industry Standard Process for Data Mining (CRISP DM) and three partitioning-based clustering techniques: K-means- PAM, Fuzzy C-Means, and a technique based on density models: Dbscan. He concludes that the planned objectives are met by identifying the same groups by the different methods applied, however, the associates of the "group obtained by PAM differ from those obtained by K-means, since the former is based on the medoid, while the latter is based on the centroid" (p. 84).

Castro and Castro (2020) develop a proposal with the objective of "identifying homogeneous and heterogeneous groups between segments based on quantitative variables of the associates in an employee fund in Colombia" (p. 10). It is supported by the Rstudio tool. They perform a principal component analysis and conclude that by applying the segmentation method with the k-means algorithm in line with the object of study, the result of this clustering determined three groups and identifies warning signs.

On the other hand, Perez (2020) presents a segmentation methodology for two Colombian financial institutions using the K-Means algorithm and comparing the results in different indexes to validate the optimal selection of the number of clusters and cluster membership. It uses the following variables: Frequency of transactions, city of origin, sector or natural person, years of seniority and transactions, with the support of R Software and the NbClustm package. It concludes that adequate segmentation is obtained.

As for background in Mexico, there is an article by Camacho et al., (2021) in which they deal with ML/FT detection by means of neural networks and an anomaly indicator. The proposed methodology involves the use of fuzzy logic to obtain risk metrics, uses four unsupervised algorithms (Strict Competitive Learning, Self-Organizing-Map, C-Means and Neural Gas) to form groups and identify the one with the highest risk, and finally applies an anomaly indicator to detect any unusual behavior. The model succeeds in reducing the false positive rate and lowering the company's costs.

It can be observed that the articles will be develo~~pe~~dmainly in Colombia with quantitative variables different from the present article, due to the requirements of the regulatory entities; most of the previous researches use data mining with the support of the CRISP DM methodology, segmentation techniques and the R software with satisfactory results.

Method

The research is non-experimental quantitative type because variables are not deliberately manipulated, but are observed as they occur in their natural context to analyze them, likewise, it is longitudinal panel type because three measurements are made over time, one for the last semester of 2022 and two semesters for the year 2023 to analyze the changes and evolution of the variables of active customers and considers a hybrid segmentation algorithm using factorial methods and Ward's method, which according to GAFILAT (2021, p. 52) is the most widely used in money laundering methodologies. 52) is the most widely used in money laundering methodologies, likewise, it is supported by data mining using the CRISP-DM methodology for the structure and process of understanding, preparation, modeling and evaluation of data, organized in:

Step 1.- Understanding the business:

The purpose of the financial institution is to contribute to economic growth by promoting the public and private sectors that generate foreign exchange in the country. These activities are carried out through the granting of credit directly or indirectly through the granting of guarantees on loans that commercial banks grant to the private sector, as well as in the financial markets through the money market desk integrated by repurchase transactions.

Step 2.- Understanding and preparing the data:

In order to ensure completeness, consistency and accuracy, the following activities are carried out:

Data Collection

Data on customers' inherent characteristics are collected by analysts and recorded in a system. The information entered in the system coincides with the information found in the physical identification file, since it is validated by the money laundering prevention area. Transactional data is obtained by extracting information from transactional applications. In order to verify the quality of the data, the external auditor's favorable opinion of the review performed for fiscal years 2022 and 2023 is considered.

Data Selection

The risk factors required by Mexican regulations that explain how and to what extent each customer represents a money laundering risk are considered:

1. Type of person: the National Risk Assessment 2020 (SHCP, 2020, p. 50) mentions that the required information may be hindered by the operation of front companies, complex legal structures, trusts and other legal agreements that allow a separation of legal ownership and the use of intermediaries, since the most complex schemes or mechanisms could allow operations with resources of illicit origin; companies and trusts are considered high risk and the rest are considered low risk.
2. Type of Politically Exposed Person (PEP): a person who performs or has performed prominent public functions in a foreign country or in Mexican territory in accordance with the provisions and who, in the case of a foreigner, is considered a high risk.
3. Seniority of the person: the typology issued by the Financial Intelligence Unit considers young persons and newly created entities as the most common mode of operation for money laundering.
4. Economic activity risk.

5. Geographical location: organized crime and drug trafficking often use the economy to launder money and bribe authorities, thus increasing crime rates and corruption in different geographical areas.
6. Product used: may increase the risk considering: The type of person with whom it can operate, the type of currency, the jurisdictions involved in its operation, the complexity of the product, the participation of financial intermediaries in its operation and whether the operation could be carried out jointly with other development banks.
7. Destination of the resource: indicates whether it is a credit in local currency (MN) or in US dollars (USD) related to the client's economic activity.
8. Origin of the resource: indicates whether it is domestic or foreign related to the client's economic activity, as these may increase the risk considering: The ease of identification of third parties involved, whether or not it allows the placement or receipt of resources in foreign currency, whether or not it allows the mobility of resources to or from abroad, and the type of process (face-to-face or non-face-to-face that gives rise to the channel's operations).
9. Number of operations: this is the sum of operations carried out in the six-month period by the type of person in the product used.
10. Amount of operations.
11. Counterparty number: indicates the number of bank accounts of different counterparties to which the funds will be deposited, considering more than five accounts as high risk.
12. Fund transfers: shows the country to which the transfer is made in order to analyze its profile and behavior.

Data Cleansing

Exploratory analysis is carried out to find and eliminate incomplete, inaccurate or incorrect records, "because hierarchical methods have no solution with missing values and outliers deform the distance and produce unit clusters" (Perez, 2013, p. 279). Table 1 below shows the number of records eliminated:

Table 1

General summary of deleted data

Customers with transaction	2022 2S	2023 1S	2023 2s
Total number of records	730	961	553
Total records deleted	3	37	19
Percentage of representation of deleted records	0.41%	3.85%	3.44%
Total clean records	**727**	**924**	**534**

Note. The table shows the total number of records and the elimination of records for inaccuracy in the heap records due to inaccuracy in the heap. The percentage of elimination is low with respect to the total number of records and therefore does not affect the result of the investigation.

Step 3.-Execution:

Phase I.- Data Exploration

A descriptive analysis was performed with the support of R software. It is identified that 53.95% are companies and 37.41% are agencies or entities; only one is assimilated foreign PEP; of the 324 economic activities, the majority are low risk: 20% Investment companies, 9% foreign financial institutions and 7% private and mixed multiple banking services, among others; the average age is 22 years, with a maximum of 123 years; the geographic location is concentrated in Mexico City and Nuevo Leon; the origin and destination of funds is mainly in Mexican pesos and 3.5% of transfers are to the United States. One to two counterparties are used to disperse resources. As for the number of operations, the minimum is one and the average is between three and four. With respect to the amount of operations, a positive bias can be observed, since the amounts recorded up to the third quartile are lower than the maximum.

Phase II.- Segmentation of the Factors Customers, Products, Channels and Geographic Location: Segmentation Technique and Identification of Factors

The classification suggested by Pardo (2020) is considered, which, based on what has been reviewed in previous research, strengthens the quantitative method in order to expect better segmentation results:

- Step 1: the corresponding principal axis analysis is performed, using on the one hand the simple correspondence method for the segmentation of products, channels and the integral segmentation of the four factors; the objective of the simple correspondence analysis is to describe the associations between the row and column variables in order to have a view of the data for its interpretation; on the other hand, the multiple factor analysis for the geographic and customer segmentation, considering that the data table is multivariate. The segmentation analysis is performed through a classification based on the coordinates of the first axes obtained from the factor analysis. These segments are formed in such a way that the elements within each segment are as similar as possible and that the elements of different segments are as different as possible. The concept of inertia or total variance of the elements to be segmented is very important in the analysis. This inertia is divided into intraclass or intragroup inertia (variance of the members of the same segment) and interclass or intergroup inertia (variance between the centers of each of the segments).
- Step 2: the number of axes for the classification is selected considering several options, on the one hand, by analyzing the variances of the main dimensions with the eig function, the eigenvalues correspond to the amount of information retained by each dimension; on the other hand, by observing the eigenvalue plot of the variances ordered from highest to lowest with the fviz_screeplot function of the Factoextra package in which visually allows to select the dimensions by observing the curvature in the bar "called elbow"; For the case of the product and channel factors, it is further supported by the quality of representation of the cosine squared rows and columns; the graphical representation of the cosine squared row and column points with the Corrplot package and the contribution of the rows and columns in order to explain the variability in the data set. In the case of the customer and geography factors, to confirm the number of dimensions, we use the Factoextra package, which provides a list of matrices containing all the results of the active variables (coordinates, correlation between variables and axes, cosine squared and contributions) and the quality of the representation of the variables in the cosine squared factor map with the Corrplot package.
- Step 3: hierarchical classification is performed using Ward's method on the "individuals" of the previous step. This method again requires the selection of a concept of similarities, dissimilarities or distances. Additionally, the selection of a distance between segments is required. To determine which individuals join first, it is necessary to calculate the matrix of distances between all pairs

of individuals. Joining the first pair results in a partition of n-1 (where n is the number of individuals to be segmented) segments, one of them with two individuals. It is required to calculate the distance between the new segment and the remaining individuals. By joining the two closest individuals we have a new distance matrix of size (n-1 ,n-1). On this matrix, the nearest pair is selected again and the algorithm continues in this way until a single segment containing all the individuals is reached.

- Step 4: the number of classes is decided and the tree is cut, using hierarchical classification. The dendrogram represents a series of embedded segments, where the number of segments decreases as the height of the tree grows. To obtain a particular number of segments, a cut is made in the tree. This tree is constructed starting from the global set of individuals (top-down classification) and making successive divisions until reaching each individual.
- Step 5: a consolidation K-means is performed starting from the centers of gravity of the partition obtained by cutting the tree.
- Step 6: the segments obtained are characterized by comparing the descriptive statistics within them with the statistics of the classified population. For continuous variables, the mean within the segment is compared with the overall mean. The profiles of each segment are accompanied by descriptive statistics (mainly the mean and standard deviation) as well as descriptive statistics of the illustrative variables of the individuals belonging to each segment.

For the customer factor, the data established by the provisions for inherent risk are as follows: Type of person, PEP, seniority of the type of person and risk of the economic activity; by its transactional characteristics: Destination and origin of the resource, number and amount of operations carried out in the six-month period, number of counterparties and country in which the funds were transferred.

For the product factor, the binary table prepared by the product development staff in conjunction with the financial institution's money laundering prevention staff and validated by the external auditor was obtained, in which they indicated with one if the product presents a risk and zero if it does not. Risk considering: The type of person with whom it can operate, the type of currency, the jurisdictions involved in its operation, the complexity of the product, the participation of financial intermediaries in its operation and whether the operation could be carried out jointly with other development banks in Mexico.

For the channel factor, the financial institution disperses and receives financial resources from its clients in five ways. The binary table previously prepared by the cash flow personnel in conjunction with the money laundering prevention personnel and validated by the external auditor was obtained, in which they indicated with one if the channel presents a risk and zero if it does not. Risk considering: The ease of identification of third parties involved, whether or not it allows the placement or receipt of resources in foreign currency, whether or not it allows the mobility of resources to or from abroad, and the type of process (face-to-face or non-face-to-face that gives rise to the channel's operations).

The segmentation of the geographic location factor is carried out considering all the states that make up Mexico, since transactions are carried out from any state. The risk indicators used for segmentation are: Homicide being the number of victims of intentional homicide per 100,000 persons; violent crime being the number of violent crimes per 100,000 persons, include: Robbery, assault, sexual violence, and violence within the family; and organized crime, which is composed of extortion, major crimes, retail drug crimes, and kidnapping or human trafficking ;(IEP, 2023).

The indicators are on a scale of one to five, with one representing the most peaceful score and five the least peaceful. Organized crime and drug trafficking often use the economy to launder money and bribe the authorities, thus increasing the crime rate and corruption and consequently the likelihood of being sentenced. On the other hand, the state competitiveness index is considered. Competitiveness is the state's capacity to generate, attract and retain talent and investment, which translates into productivity and well-being for its inhabitants and investors.

Phase III.- Risk Measurement and Classification Considering an Abnormality Indicator:

The risk level indicates how critical the customer is considering the different threats involved in each factor. To measure it, a matrix is prepared with the result of the segmentation, in which a value of one is considered for low risk, two for medium risk and three for high risk.

Table 2 shows the results in which it is observed that the factors: Customers are at low and medium risk levels; all products are classified as medium risk; channels and geographic location have high and medium risk levels.

Table 2

Customer Risk Grades by semester

Factor	Period 2022 second semester	Period 2023 first semester	Period 2023 second half
Customer segment 1	2	1	1
Customer segment 2	1	2	2
Customer segment 3	2	1	2
Customer segment 4	2	1	1
Product segment 1	2	2	2
Product segment 2	2	2	2
Product segment 3	2	2	2
Product segment 4	2	2	2
Channels segment 1	2	2	2
Channels segment 2	3	3	3
Channels segment 3	2	2	2
U. Geographic segment 1	2	2	2
U. Geographic segment 2	3	3	3
U. Geographic segment 3	3	3	3
U. Geographic segment 4	2	2	2

Risk classification is fundamental to effectively allocate the financial institution's available resources to mitigate ML/FT risk. In order to classify the client by the degree of risk in each six-month period, the sum of the value assigned in each risk factor is added, so that a high risk level has an upper limit of 12 and a lower limit of 11; for medium-high risk the upper limit is 10 with a lower limit of nine; medium risk has an upper limit of eight and a lower limit of five; and for low risk it is equal to 4. Table 3 shows a summary of the results showing a higher concentration of medium-risk clients.

Table 3
Concentration of customers in risk grades by semester

Period	Risk under	Risk medium	Risk medium high	Risk high
2022 second half	0	326	401	0
2023 first half	0	850	74	0
2023 second half	0	437	97	0

Segmentation integrates homogeneous individuals allowing the identification of atypical individuals within the segment. To analyze abnormal behavior it is common to use in risk management two standard deviations of the average amount traded; therefore, it is applied in the research considering that it is the one used in the current model of the financial institution. An integral segmentation of the four factors is carried out, creating groups that are homogeneous inside and heterogeneous outside, in order to subsequently generate a table with the operations of maximum amount, average amount and average amount plus two standard deviations per client in the three periods to compare the amount of the average operation plus two standard deviations and the maximum operation carried out by the client with that of the segment and thus be in a position to suspect an unusual operation.

Results

For the customer factor, the result is five segments for 2022, four for the first and second half of 2023, as an example, for the second half of 2023, Figure 1 shows the first factorial plane in which the conformation of four associated groups can be observed and Figure 2 shows the dendrogram with four segments according to the analysis performed together with the eigenvalues/variance of the principal components, the values plot with the elbow technique and the correlation plot of variables. The Ttable 4 shows the result of the hierarchical classification with Ward's method and optimization of the classes with K-means.

Figure 1
Graphical representation of data in two dimensions

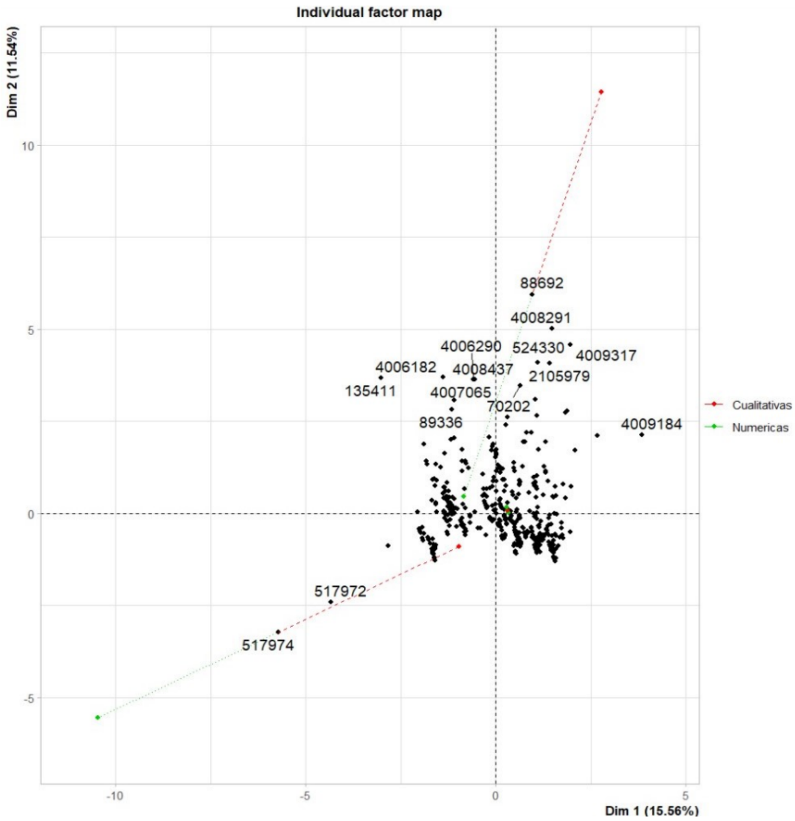


Figure 2
Graphical representation of the dendrogram

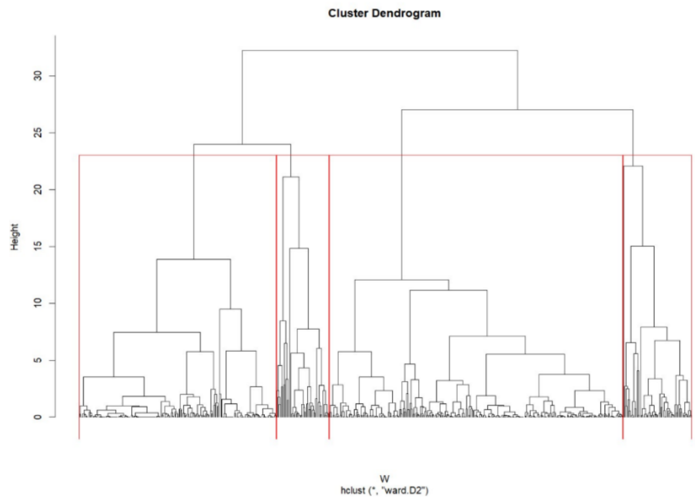


Table 4
Hierarchical classification with Ward's method and optimization with K-means

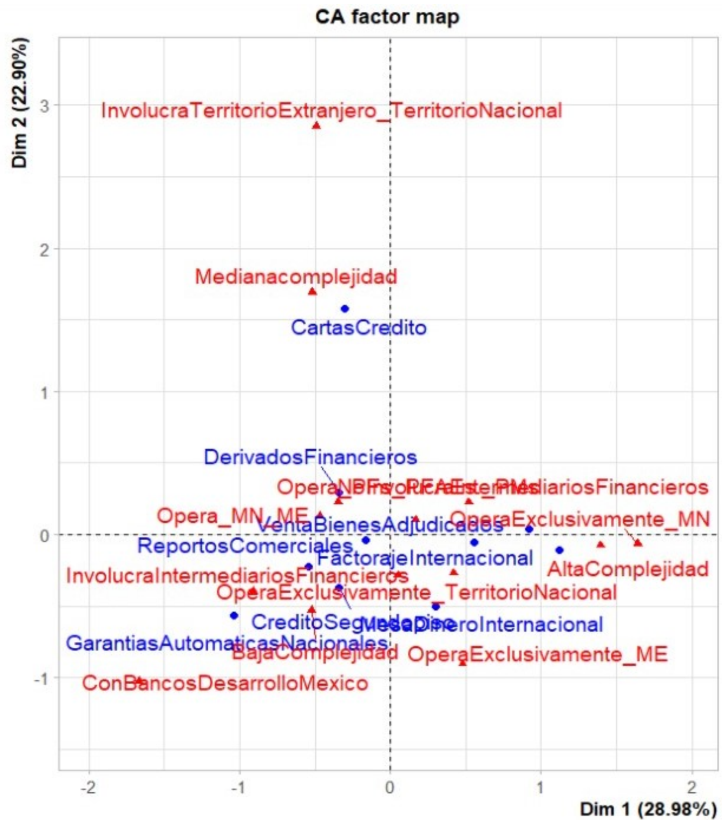
Method	Segment one	Segments two	Segment three	Segment Four
Ward	256	60	46	172
K-means	248	59	39	188

The clients that are grouped in the different segments in the three semesters are similar inside and different outside, this is corroborated by the descriptive statistics that, for example, in the first segment groups 46.44% of the total number of clients in the second semester of 2023 95.17% are companies with 57% of the origin of the national resource and 100% of its destination in national currency; the risk of the economic activity is medium; the minimum number of operation is one and its maximum 23; the average amount of its operations goes from $1 to $ 3,949,492,386 (MN) and the third segment with 35.21% of the clients being 91.44% dependencies and entities with 68% of the origin of the foreign resource and 100% of its destination in national currency; the risk of the economic activity is low; its minimum number of operation is one, but its maximum is 14 and the amount of its operations goes from $5,000 to $286,160,000,000 (MN). Customers in the second segment have almost similar characteristics to those of the first segment, differing mainly in the number and amount of transactions; the same is true for the third segment and the fourth segment.

For the segmentation of the products, the factorial plane in Figure 3 shows that letters of credit have different risks with respect to the others; as well as, that of domestic automatic guarantees and international money desk, the quality of the rows is obtained in which it turns out that nine of the eleven products are well represented up to the fifth dimension by bringing their sum close to one and only two: second floor credit and commercial repo are far away; the row dot plot in five dimensions of the cosine squared which shows that their highest values are found within the first four dimensions and the table of the contribution of the product rows in percentage, which shows the products with the highest value and which are the ones that contribute most to the definition of the dimensions, being in the first dimension trust services and domestic automatic guarantees and in the second dimension letters of credit; as for the table of the quality of the columns, which shows whether a risk is well represented by the five dimensions if the sum is close to one, it turns out that 12 out of 14 are well represented and only two are well represented: operates in MN and ME; and operates exclusively ME are little far from one and with the plot of the column points in five dimensions of the cosine squared showing all the products in which it can be observed that their highest values are within the first three dimensions and the result by the mixed algorithm are five segments obtained in accordance with what was observed in the first factorial plane and in the histogram together with the eigenvalues/variance of the main dimensions and with the elbow technique.

Figure 3

Factorial foreground of the attributes of the product variable

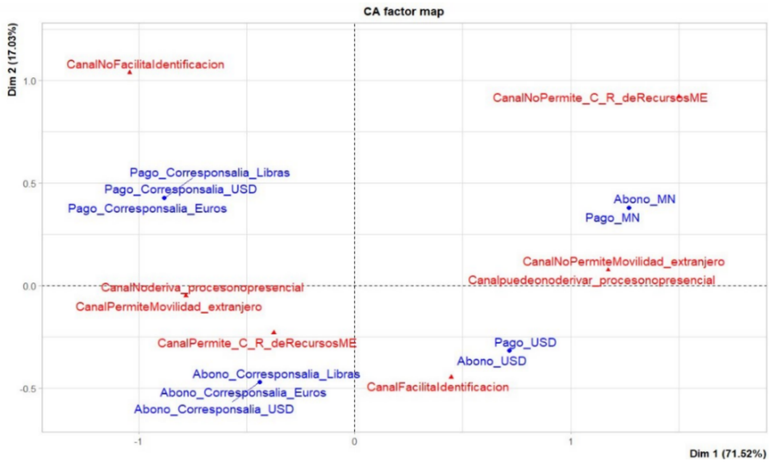


With the result of the hierarchical classification with Ward's method and optimization of the classes with K-means, the products contained in two of the segments are as similar as possible, differentiating them from the three products that formed one segment each based on the analysis performed. The possibility of forming four segments is evaluated in which it is observed that the product of national automatic guarantees is part of the first segment, affecting the quality of the rows.

For channel segmentation, the factorial plane in Figure 4 shows that correspondent payment channels in pounds, Euros and USD do not facilitate identification, that payments and credits in MN do not allow mobility abroad and are not face-to-face, and that correspondent payments in pounds, Euros and USD and payments and credits in USD facilitate customer identification. The table of the cosine squared of the rows channels is obtained, which shows whether a product is well represented by the three dimensions if the sum is close to one, resulting that all are well represented, also, the graph of the row points in five dimensions of the cosine squared that shows all the products within the three dimensions, the table of the contribution of the row channels in percentage, which shows the channels with the highest value and which are the ones that contribute the most to the definition of the dimensions, being in the first dimension subscription and payment in MN and in the second dimension subscriptions in correspondent; the quality of the columns is obtained using the same techniques.

Figure 4

Factorial foreground of the attributes of the carcass variable

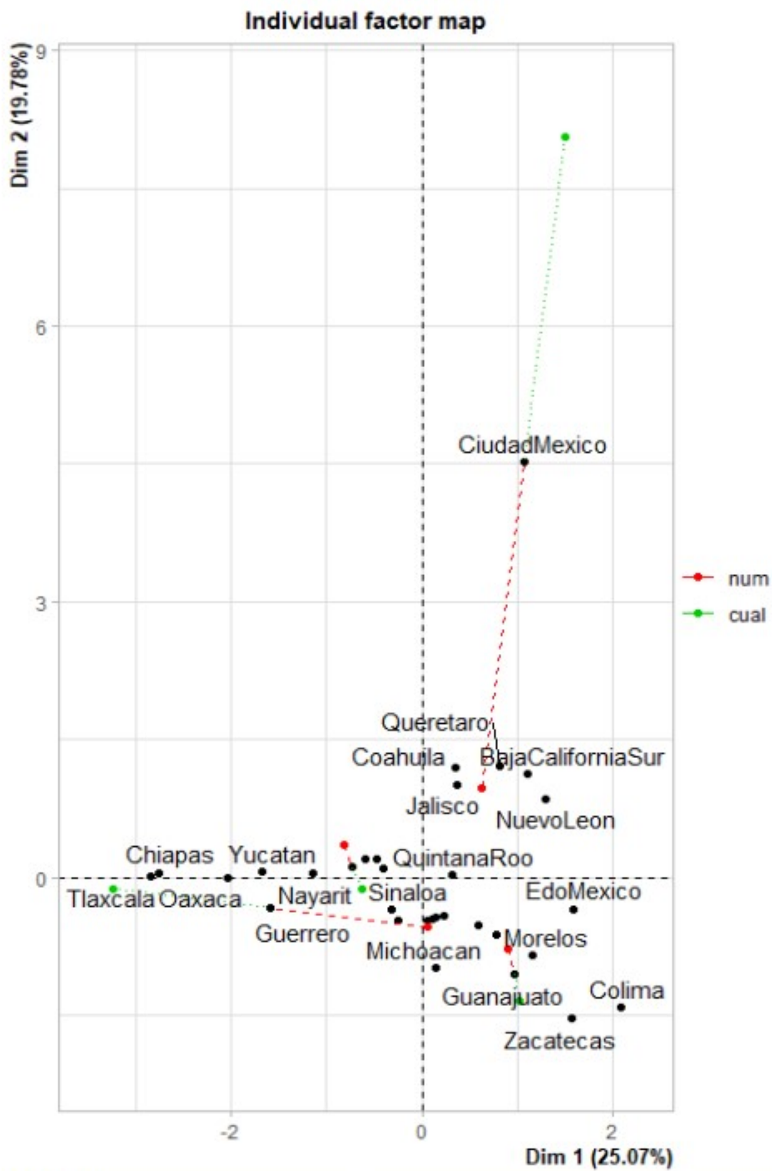


With Ward's method and K-means class optimization, the channels contained in the three segments are as similar as possible on the inside and differentiated on the outside based on the analysis performed. The possibility of forming two segments is evaluated in which the first segment consists of payments and credits in MN and USD and the second segment consists of payments and credits in correspondent, the latter is divided into two in the proposal; observing the symmetrical diagram of the factorial plane, the payment and credit in USD are closer to the credit in correspondent, so it would be assumed different groupings to those obtained, but analyzing the contribution and quality of the rows and columns in three dimensions it is concluded that it is the best option.

The Figure 5 shows the factorial plane of the geographic location that allows us to observe the conformation of four groups, which were corroborated by obtaining the eigenvalues/variance, the graph of values ordered from highest to lowest and the graph of points in the cosine squared dimensions which shows that the highest values of the numerical variables are found in the first and even the fourth dimension for the qualitative variable and with the result of the hierarchical classification with Ward's method and optimization of the classes with K-means the geographic locations contained in the four segments are as similar as possible in their interior differentiating themselves in their exterior based on the analysis carried out. The highest risk segments are the second, comprising the states of Baja California Sur, Coahuila, Jalisco, Nuevo León and Querétaro, which are characterized by a high competitiveness index; and the third segment, comprising the states of Campeche, Colima, Guanajuato, Hidalgo, Michoacán, Morelos, Puebla, San Luis Potosí, Estado de México, Tabasco, Veracruz and Zacatecas. They are characterized by having a medium-low competitiveness index and this is supposed to be caused by being one of the states with the highest perception of crime and insecurity; it integrates Mexico City with a very high competitiveness index.

Figure 5

Factorial foreground of the attributes of the variable geographic location



To conclude, the segmentation of the four factors for each semester is integrated by applying a correspondence analysis Figure 6 shows the factorial plane obtained in which the concentration of the cloud of points in four groups can be observed, the blue points and numbers represent the clients, while the red ones represent the location of the risk factors. These results were validated with the histogram were validated with the histogram together with the eigenvalues/variances of the main dimensions and their plot with the elbow technique and with the result of the hierarchical classification with Ward's method and optimization of the classes with K-means.

Figure 6
Factorial close-up of the four segments in the second half of 2023

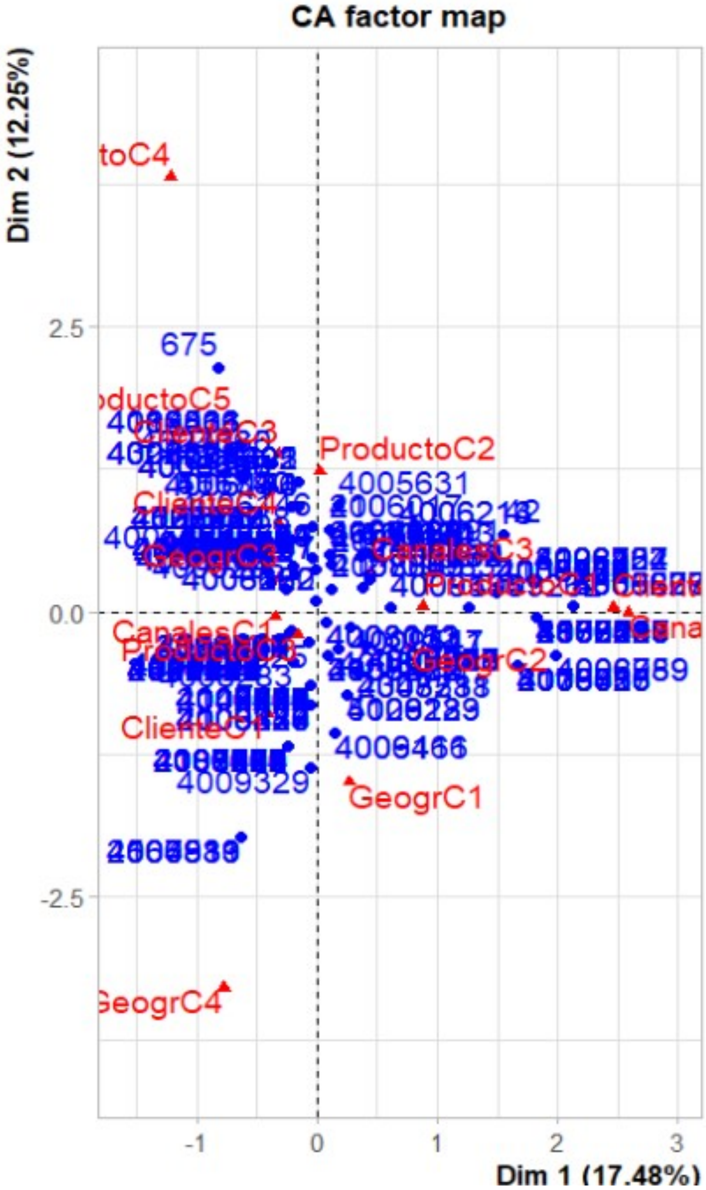


Table 5 shows the results of the clients that presented an alert shows the results of the clients that presented an alert and that considering the grouped amounts of the segmentation, a threshold of the maximum amount of operations is established in comparison with those of the client in that period and previous periods, which allows to know the client's transactionality and to accept or reject the alert and thus reduce false positives.

Table 5
Clients with alerts

Period	Total	With alert	Confirm the alert	Medium risk alert	Alert medium high risk
2022 second half	727	1	0	0	1
2023 first half	924	518	14	13	1
2023 second half	534	183	15	6	9

Discussion and Conclusions

In conclusion, Mexico, according to BASEL, faces a medium risk situation in relation to ML/FT and, therefore, requires strengthening its prevention, control and detection strategies.

For criminal organizations, financial institutions are useful and of interest, since their international connection allows them to operate large amounts of money throughout the world and thus facilitate their actions, hence the importance and responsibility of financial institutions to act in an effective and timely manner in their preventive management (Guevara and Flores, 2021, p. 5).

The objective and research questions posed are met, since the segmentation technique suggested by Pardo (2020) allows grouping, explaining and validating the homogeneous integration inside and heterogeneous integration outside, as well as to measure and explain the degree of final customer risk. With the data collected during the three periods together with the values grouped in each segment, it is possible to establish a threshold and buy the transactionality of the customers in order to have more knowledge, comparability and elements to determine the abnormality of the operations and thus decide in a better manner unusual transactions with a risk-based approach and perform a more specific knowledge management of customers.

The results of the proposed segmentation method reflect a medium and high risk profile of the financial institution's clients in terms of money laundering, as opposed to the method currently used, which concentrates on low risk. This is explained by Mexico's geographic position and the risks derived from its proximity to other jurisdictions (Ministry of Finance and Public Credit, 2020) and the indicators taken from IEP (2023), as well as by the risks grouped in the factors of products and channels used together with the inherent and transactional characteristics of the customers.

This article proposes a model that considers the recommendations described in FATF reports, best practices and national provisions; as well as variables that had not been included in previous studies, assigning a degree of risk and thresholds for the identification of suspicious transactions and therefore improving it.

One of the limitations of the model is the dependence on data quality, so it is recommended to carry out the cleaning process and; establish and adequately maintain the quality of the data to be in a position to prevent the proper functioning of the Method, likewise, is the absence of data from confirmed clients in money laundering to know and validate the behavior of the model.

It is worth recalling the BASEL (2020) guidelines in which it is essential that financial institutions understand and document the normal and expected banking activity of their customers. This will make it possible to establish effective mechanisms to detect operations that deviate from the usual pattern of banking activity.

For future research, it is suggested that indicators of the country's economic activity be incorporated to compare it with the behavior of the client and the segment and to assess whether the increases or decreases are consistent with each other in order to improve the method.

References

Basel Institute on Governance. (2020). *Sound management of risks related to money laundering and financing of terrorism: revisions to supervisory cooperation [Gestión adecuada de los riesgos relacionados con el blanqueo de capitales y la financiación del terrorismo: revisiones de la cooperación supervisora].* https://www.bis.org/bcbs/publ/d505.htm

Basel Institute on Governance (BASEL). (2022). *Basel AML Index 2022: 11th Public Edition - Ranking money laundering and terrorist financing risks around the world.* https://baselgovernance.org/publications/basel-aml-index-2022

Basel Institute on Governance (BASEL). (2023). *Basel AML Index 2023: 12th Public Edition - Ranking money laundering and terrorist financing risks around the world.* https://baselgovernance.org/publications/basel-aml-index-2023

Camacho, M., Rocha, J., & Segovia, M. (2021). Money laundering and terrorism financing detection using neural networks and an abnormality indicator. *Expert Systems with Applications, 169, 114470.* https://doi.org/10.1016/j.eswa.2020.114470

Castro, N., & Castro, M. (2020). *Segmentación de clientes en un fondo de empleados para identificar los riesgos de lavado de activos y financiación del terrorismo*. https://repository.libertadores.edu.co/items/04da680f-e224-4b18-a09f-9cce5ebc0706

Comisión Nacional Bancaria y de Valores (CNBV). (2020). *Guía para el Fortalecimiento del Régimen Preventivo de Lavado de Dinero y Financiamiento al Terrorismo, derivada de los Hallazgos recurrentes detectados durante los procesos de Supervisión: Apartado 3. Conocimiento del Cliente y Debida Diligencia Reforzada.* https://issuu.com/gmc360/docs/2020_07_15_gui_a_hallazgos_recurrentes

Grupo de Acción Financiera de Latinoamérica (GAFILAT). (2021). *Análisis estratégico sobre el uso de Inteligencia Artificial, Minería de Datos y Análisis de Big Data en prevención y detección LA/FT (UIF/MP).* https://biblioteca.gafilat.org/?cat=17

Grupo de Acción Financiera de Latinoamérica (GAFILAT). (2023). https://www.gafilat.org/index.php/es/inicio

Guevara Rabanal, M. L., y Flores Tananta, D. C. A. (2021). Lavado de activos y su relación con la rentabilidad, banco de la nación, 2020. *Ciencia Latina Revista Científica Multidisciplinar, 5*(5), 8212-8227. https://doi.org/10.37811/cl_rcm.v5i5.899

Institute for Economics & Peace (IEP). (2023). *Mexico Peace Index 2023: Identifying and Measuring the Factors That Drive Peace, Sydney, May 2023*. http://visionofhumanity.org/resources

Jovel, T. (2020). *Desarrollo de un modelo analítico para la segmentación de asociados en una cooperativa de ahorros y crédito*. [Tesis Maestría, Universidad Nacional de Colombia]. https://repositorio.unal.edu.co/handle/unal/79648

Martínez, G., Álvarez, Y. I., y Silva, H. (2022). Empresas y prevención del lavado de dinero en México. *3C Empresa. Investigación y pensamiento crítico, 11*(1), 67-83. https://doi.org/10.17993/3cemp.2022.110149.67-83

Pardo, C. (2020). *Estadística descriptiva multivariada. Universidad Nacional de Colombia*. https://repositorio.unal.edu.co/handle/unal/79914

Pérez, L. (2013). *Técnicas de Segmentación. Conceptos, herramientas y aplicaciones*. Alfaomega Grupo Editor, S.A. de C.V

Perez, L. (2020). *Metodología para segmentación de un SARLAFT*. https://repositorio.unbosque.edu.co/bitstream/handle/20.500.12495/6168/Perez_Perez_Lincoln_Ernesto_2020.pdf?sequence=2&isAllowed=y

Secretaría de Hacienda y Crédito Público. (2020). *Evaluación Nacional de Riesgos. Versión Pública del mes de septiembre de 2020.* https://www.pld.hacienda.gob.mx/work/models/PLD/documentos/enr2020.pdf

Secretaría de Hacienda y Crédito Público (SHCP). (2022). *Disposiciones de Carácter General a que se refiere el Artículo 115 de la Ley de Instituciones de Crédito. Última Reforma publicada en el Diario Oficial de la Federación el 03 de marzo de 2022.* https://www.gob.mx/cms/uploads/attachment/file/709513/DCG_Compiladas_Instituciones_de_Credito_08.03.2022.pdf